IMPLICIT NEURAL REPRESENTATION FOR HYPERSPECTRAL VIDEO COMPRESSION

Alfredo Scalera, Paul Murray, Jaime Zabalza

University of Strathclyde
Department of Electronic
& Electrical Engineering
Glasgow, UK

ABSTRACT

With the advent of snapshot cameras, hyperspectral video is becoming more readily available. In recent years, new applications have emerged which have led to increasingly larger datasets. However, hyperspectral video compression remains in the early stages. In this study, we explore the use of implicit neural representation as a candidate solution. We propose a novel extension of an existing RGB video compression model, achieving Bjøntegaard Delta PSNR gains of $+4.99$ dB and Bjøntegaard Delta rate of -88.88% compared to traditional hyperspectral image compression methods applied frame-by-frame. In addition to reconstruction quality, the effects on downstream task performance are measured in the form of object tracking success. Compared to video compressed with methods based on principal component analysis and JPEG2000 in low data regimes, our proposed method improves tracking area under the curve by up to 23.42% and distance precision by up to 35.56% on examples from the HOT2026 dataset.

***Index Terms*—** Hyperspectral video, Video compression, Implicit neural representation, Computer vision, Dimensionality reduction

1. INTRODUCTION

Hyperspectral (HS) imagery is commonly used in: quality assessment of food; agriculture; remote sensing; and medical analysis, among others [1]. Through technological innovations in camera sensors, HS videos have become more prevalent in recent years. The advent of snapshot hyperspectral sensors has made a number of varied applications emerge, from HS object tracking [2] to scene segmentation [3]. Yet, the field dedicated to HS video compression is still in its infancy and, so far, image compression methods for HS data have been mainly applied frame-by-frame [4].

In this paper a more wholistic approach is taken towards compressing HS video. Consequently, we propose the use of implicit neural representation (INR) for HS video compression [5]. INR-based methods leverage the strong data representations enabled by neural networks while avoiding the need for large datasets. The main contributions in this paper are: Novel extension of INR-based compression method to HS; Comparison against common PCA-based approach for HS imagery; Quantifying gains from the INR-based method compared to current practices.

Author email addresses: alfredo.scalera.2019@uni.strath.ac.uk; {paul.murray, j.zabalza}@strath.ac.uk

2. RELATED WORK

Video compression in the domain of RGB video is a mature field in both academic and industrial settings. Most video systems today are built on top of standardised classical hybrid codecs such as AVC [6], HEVC [7], and VVC [8] which have been developed in the last 25 years.

Over the past decade, with the advent of deep learning, research and development in learned video codecs has emerged. These can broadly be split into two main categories: general and overfit video codecs. General learned video codecs search compressing functions that are learned from end-to-end training on large amounts of data [9], [10]. Overfit video codecs, on the other hand, leverage the overfitting abilities of neural networks thus obtaining an approximation of the functions that generate each pixel or frame of a video. Examples include HNeRV [5], NVRC [11], and COOL-CHIC [12].

In recent years, snapshot HS video cameras have emerged due to innovations in mosaic-array sensors [13]. This has led to the collection and release of various HS video datasets such as HS object tracking (HOT) challenge (HOT2020 [2], HOT2026 [14], IMEC25 [13]) and automated driving systems (HSI-Drive [3]).

HS video compression is often achieved by compressing each datacube frame individually, making use of the spectral correlation between channels [15], [16]. The work in [4] uses sparse tucker tensor decompositions per-frame, outperforming other HS image compression methods applied in the same manner. Authors in [17] propose an HS video coder by building on top of an HS image coder to compensate both spectrally and temporally, achieving higher compression rates

than applying only spectral prediction.

To the best of our knowledge, overfit neural compression methods are yet to be applied to the field of HS video compression. In this work we explore their application to the problem and specifically target INR to show the potential of overfit compression.

3. METHODS

In the following sections, we detail the methods employed in this study, as well as the experimental setup.

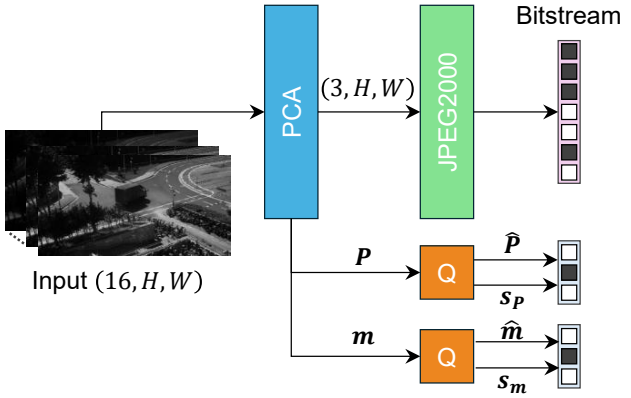


Fig. 1. HS video compression with PCA+JPEG2000. The projection matrix $\boldsymbol{P}$ is quantized to $\hat{\boldsymbol{P}}$ before being written into the bitstream. Q represents min-max quantization and $\boldsymbol{s_P} = (\boldsymbol{P}_{\min}, \boldsymbol{P}_{\text{scale}})$ and analogously for $\boldsymbol{m}$.

3.1. PCA-based Video Compression

PCA is commonly used for spectral dimensionality reduction, and its combination with RGB compression standards has been proposed for HS image and video compression.

Here we explore PCA and JPEG2000 as first proposed in [18]. The technique is chosen as a strong baseline for overfit HS video compression (since the principal components must be fit to each frame or video). Furthermore, both PCA and JPEG2000 are familiar to the hyperspectral community and readily deployable.

PCA is applied individually to each frame and linearly projects the input datacube $\boldsymbol{D} \in \mathbb{R}^{C_{\text{in}} \times H \times W}$ from C_{in} channels to $C_{\text{out}} = 3$ channels to fit JPEG2000. The reduced datacube is then compressed with JPEG2000 to the desired target bits-per-pixel (bpp). The PCA projection matrix $\boldsymbol{P} \in \mathbb{R}^{C_{\text{in}} \times C_{\text{out}}}$ and mean vector $\boldsymbol{m} \in \mathbb{R}^{C_{\text{in}}}$ are min-max quantized to improve memory footprint (Eq. 1).

$$q = \left\lfloor Q_{\max} \frac{\boldsymbol{x} - \boldsymbol{x}_{\min}}{\boldsymbol{x}_{\text{scale}}} \right\rceil \tag{1}$$

Where x represents the full-precision value, q is the corresponding quantized value, $Q_{\max}$ the maximum of the quantized range, and $\lfloor \cdot \rceil$ is the rounding operation. Symmetric signed min-max quantization requires storing the minimum values $\boldsymbol{x}_{\min} = \min(\boldsymbol{x})$ the scale factors $\boldsymbol{x}_{\text{scale}} = \max(\boldsymbol{x}) - \min(\boldsymbol{x})$ for dequantization. These are included in the JPEG2000 bitstream as side information $\boldsymbol{s_P} = (\boldsymbol{P}_{\min}, \boldsymbol{P}_{\text{scale}},)$ and $\boldsymbol{s_m} = (\boldsymbol{m}_{\min}, \boldsymbol{m}_{\text{scale}},)$.

When decoding the inverse operations are applied: first the JPEG2000 bitstream is decoded, the projection matrix and mean vector are de-quantized, and finally inverse PCA is applied to the reduced datacube to restore all C_{in} channels, obtaining the decompressed datacube $\tilde{\boldsymbol{D}}$.

3.2. INR-based Video Compression

Implicit neural compression can be leveraged in various ways for video compression. Forms include compressed models that learn to map pixel or frame coordinates to pixel values or frame tensors; hybrid schemes where the latent variables and decoders of autoencoder-like networks are compressed [5]; and fully end-to-end overfit video codecs [11], [12].

In this paper, we adapt HNeRV [5] for HS videos. HNeRV is a hybrid RGB approach, treating video compression as a model and latent compression problem (as shown in Fig. 2).

We apply three extensions to HNeRV: the encoder and decoder networks are adapted to allow multiple channel inputs, $L2$ loss is replaced with bespoke loss HSFusion, and a hyperparameter search was conducted to find a more optimal combination for the HOT2026 dataset. We refer to the hyperspectral version as HS-HNeRV

The HSFusion loss (Eq. 2) combines L2 and cosine-similarity losses. The former targets general reconstruction, while the latter ensures higher spectral fidelity in each pixel.

$$\text{HSFusion} = \lambda \cdot \text{L2}\left(\boldsymbol{D}, \tilde{\boldsymbol{D}}\right) + (1 - \lambda) \cdot \text{Sim}\left(\boldsymbol{D}, \tilde{\boldsymbol{D}}\right) \tag{2}$$

Here $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$ are the original and reconstructed frame datacubes, respectively. Sim is cosine-similarity, while $\lambda \in [0, 1]$ weighs the two losses.

3.3. Experiments

We measure the effectiveness of INR compared to PCA+JPEG2000 for HS video compression in two ways: (1) reconstruction quality metrics and (2) downstream task performance.

PSNR and spectral angle mapper (SAM) are applied to measure the fidelity of the decompressed representations. The first measures general reconstruction, while the latter evaluates the spectral divergence of decoded pixels.

Hyperspectral object tracking (HOT) is chosen as a representative downstream task to measure the impact of compression on computer vision algorithms. HOT consists of HS

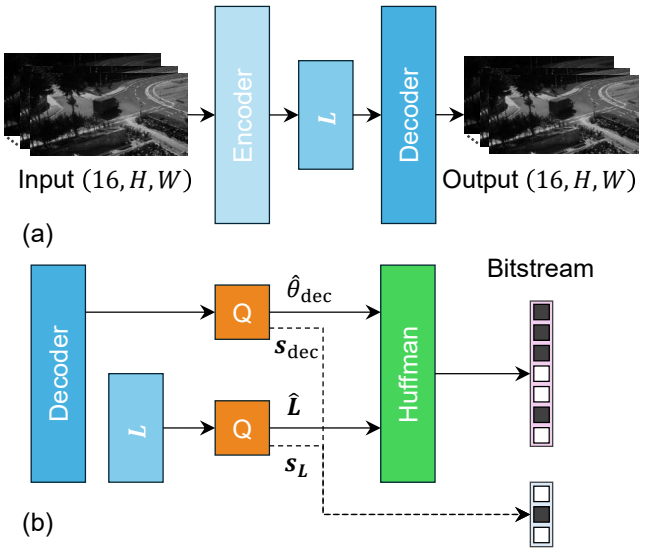


Fig. 2. HS video compression with HNeRV-based method. (a) High-level frame reconstruction for overfitting. (b) Compression - decoder parameters θ_{dec} are shared across all video frames while each frame has a unique set of latent variables $\boldsymbol{L}$. Huffman coding losslessly compresses quantized latents $\hat{\boldsymbol{L}}$ and decoder parameters $\hat{\theta}_{\text{dec}}$.

videos with moving targets, with the bounding box of a single target provided in the first frame. Algorithms track the object by temporally propagating the bounding box through each frame. We evaluate CSSTrack [19], which represents the current state-of-the-art for HOT tasks. AUC and DP are measured on both raw and decompressed sequences. AUC is the area under the curve of the success plot, which measures the percentage of frames where the intersection-over-union exceeds a particular threshold. DP is distance precision measured as the proportion of frames in which the bounding box center is below a distance error of 20 pixels.

Four test video sequences were obtained from the HOT2026 [14] VIS training set [1]. The sequences have a native resolution of 512×272 pixels and 16 spectral bands and were center cropped to 512×256 in all experiments (achieving a 2:1 aspect ratio to accommodate the model).

HS-HNeRV model sizes and PCA+JPEG2000 target bpp are swept through and PSNR and SAM metrics are computed between the decompressed and original frames for each test sequence. To obtain per-video values, both PSNR and SAM are averaged across all frames.

For performance on HOT tasks, AUC and DP are obtained over three HS-HNeRV model sizes and on PCA+JPEG2000 sequences with matching bpp. We use a CSSTrack checkpoint trained on HOT2020[2].

[1] `heartsurgery5, high_playground1, high_truck, snowcat`

[2] `https://github.com/hscv/CSSTrack`

PCA projection matrix values are quantized to 8-bit unsigned integers after fitting, with minimum and scale factors (side information) embedded as 16-bit floats in the bitstream, as illustrated in Fig. 1.

As for HS-HNeRV, the same hyperparameters are set for each video: 64 encoder channels, 16 latent channels, and a learning rate of $1.1\text{e}{-3}$ and trained for 300 epochs with $\lambda = 0.7$. These were obtained via a random hyperparameter search with 15 trials on four videos from the HOT2026 validation set [3], maximizing peak signal-to-noise ratio (PSNR). To obtain different bpp values, the model size was varied (3.0, 1.5, 0.75, 0.375 million parameters, respectively).

4. RESULTS

Rate-distortion curves are provided in Fig. 3. Four HS-HNeRV sizes and six PCA+JPEG2000 bpp targets were evaluated. Values are averaged across all four test sequences. Over equivalent bpp ranges, HS-HNeRV achieves both higher PSNR and lower SAM scores. Specifically, HS-HNeRV achieves a Bjøntegaard Delta (BD)-rate of -88.88% and a BD-PSNR of 4.99 dB with respect to PCA+JPEG2000.

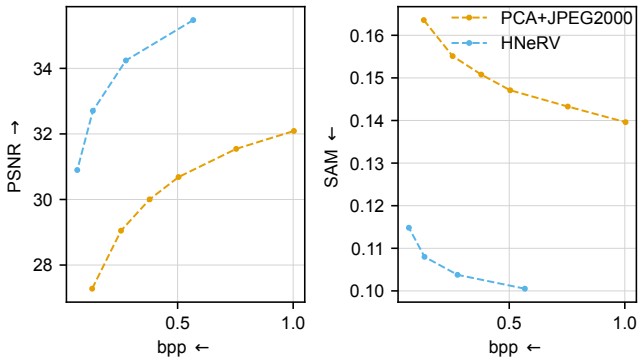


Fig. 3. INR-based compression affords higher fidelity. HS-HNeRV surpasses PCA+JPEG2000 in both spatial (left) and spectral (right) reconstruction at similar compression levels.

In low bitrate regimes, HS-HNeRV also outperforms PCA+JPEG2000 in maintaining tracking performance, as reported in Table 1. For bpp of approx. 0.13, HS-HNeRV improves over the former by 23.42% and 35.56% in tracking AUC and DP, respectively. On the other hand, PCA+JPEG2000 surpasses HS-HNeRV at lower compression rates. Interestingly, the DP of CSSTrack improves on sequences compressed with PCA+JPEG2000 at 0.6 bpp compared to the raw data.

[3] `athlete1, car8, cat2, pingpong4`

Table 1. CSSTrack performance on compressed HOT2026.

Compression Method	bpp	AUC	DP
Raw Sequence	–	43.59	67.84
PCA+JPEG2000	0.135	24.55	42.80
	0.275	27.87	46.44
	0.604	37.92	69.92
HS-HNeRV	0.133	30.30	58.02
	0.276	33.37	61.24
	0.568	35.02	62.48

5. DISCUSSION

INR-based methods show clear potential for HS video compression compared to PCA-based overfit approaches. Stronger representation is seen across all bitrates with HS-HNeRV (example illustrated in Fig. 5), achieving the same PSNR as PCA+JPEG2000 while requiring only 11.12% of the data (BD-rate of -88.88%). The potential for low-data environments is also demonstrated by the increased HOT performance. Overall, an INR-based overfit approach removes the need for large amounts of data for a general model while leveraging the function fitting capabilities of deep learning.

The proposed HSFusion loss is also beneficial for algorithms that operate at the pixel-level as it forces spectral consistency compared to the standard L2 loss, as seen in Fig. 4.

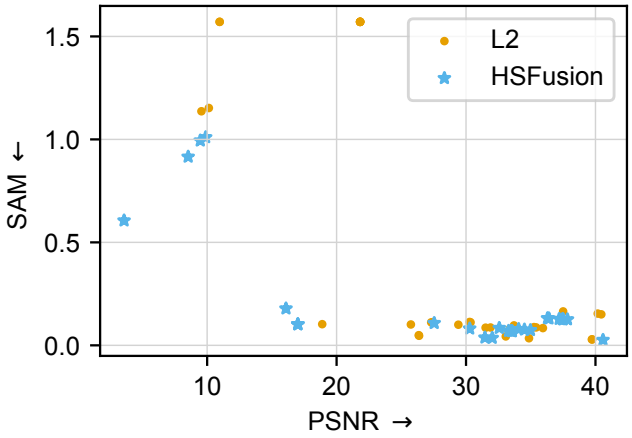
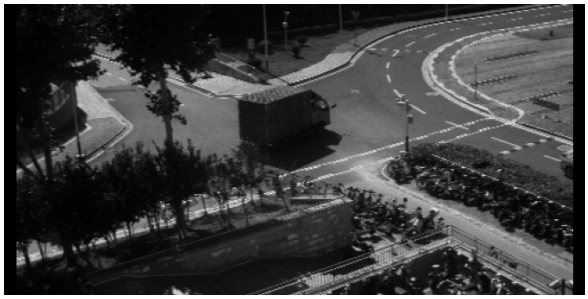


Fig. 4. Generally, for equivalent PSNR values, HSFusion forces better spectral fidelity within each pixel (SAM metric). Each datum represents a hyperparameter search trial.

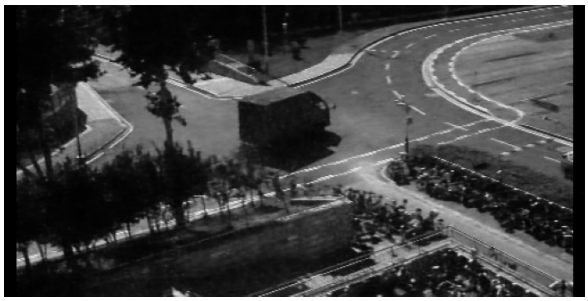
Higher compression ratios could be further achieved by employing more advanced quantization and encoding schemes. For example, quantization-aware training and compact entropy models. The performance of HNeRV and similar methods is highly dependent on the right choice of hyperparameters. General hyperparameters can be found via search strategies on multiple videos, but this defeats the purpose of dataset-free compression when input distributions differ. This can be mitigated by adding hyperparameter optimization strategies, such as Bayesian or gradient-based optimization.



(a) Ground Truth



(b) PCA+JPEG2000 @0.604bpp



(c) HS-HNeRV @0.568bpp

Fig. 5. Qualitative comparison of reconstruction capabilities. Channel 16 of first frame from `high_truck` sequence. HS-HNeRV achieves a higher spatial and spectral quality.

6. CONCLUSION

In this work, we demonstrate the way in which INR can be leveraged for HS video compression. The methods shown demonstrate strong performance compared to PCA combined with JPEG2000 for video encoding, improving reconstruction quality by $+4.99$ dB.

Future work will include the adoption of bespoke architectures for HS video, the use of hyperparameter optimization for ideal overfitting, and adjusting quantization and encoding procedures to match the latest overfitting RGB codecs.

7. REFERENCES

[1] A. Bhargava, A. Sachdeva, K. Sharma, M. H. Alsharif, P. Uthansakul, and M. Uthansakul, "Hyperspectral imaging and its applications: A review," *Heliyon*, vol. 10, no. 12, e33208, Jun. 30, 2024, ISSN: 2405-8440. DOI: 10.1016/j.heliyon.2024.e33208.

[2] F. Xiong, J. Zhou, and Y. Qian, "Material Based Object Tracking in Hyperspectral Videos," *IEEE Transactions on Image Processing*, vol. 29, pp. 3719–3733, 2020, ISSN: 1941-0042. DOI: 10.1109/TIP.2020.2965302.

[3] K. Basterretxea, V. Martínez, J. Echanobe, J. Gutiérrez–Zaballa, and I. Del Campo, "HSI-Drive: A Dataset for the Research of Hyperspectral Image Processing Applied to Autonomous Driving Systems," in *2021 IEEE Intelligent Vehicles Symposium (IV)*, Jul. 2021, pp. 866–873. DOI: 10.1109/IV48863.2021.9575298.

[4] S. Das, "Hyperspectral image, video compression using sparse tucker tensor decomposition," *IET Image Processing*, vol. 15, no. 4, pp. 964–973, Mar. 1, 2021, ISSN: 1751-9667. DOI: 10.1049/ipr2.12077.

[5] H. Chen, M. Gwilliam, S.-N. Lim, and A. Shrivastava. "HNeRV: A Hybrid Neural Representation for Videos," Accessed: Jun. 3, 2026. [Online]. Available: https://arxiv.org/abs/2304.02633v1, pre-published.

[6] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003, ISSN: 1558-2205. DOI: 10.1109/TCSVT.2003.815165.

[7] N. Minallah, S. Gul, and M. Bokhari, "Performance Analysis of H.265/HEVC (High-Efficiency Video Coding) with Reference to Other Codecs," in *2015 13th International Conference on Frontiers of Information Technology (FIT)*, Dec. 2015, pp. 216–221. DOI: 10.1109/FIT.2015.46.

[8] B. Bross et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021, ISSN: 1558-2205. DOI: 10.1109/TCSVT.2021.3101953.

[9] Z. Jia et al. "Towards Practical Real-Time Neural Video Compression," Accessed: Jun. 3, 2026. [Online]. Available: https://arxiv.org/abs/2502.20762v2, pre-published.

[10] H. He, Y. Shi, J. Wang, J. Liang, N. Ling, and B. Yin, "End-to-end learned video compression: A comprehensive review," *Neurocomputing*, vol. 694, p. 133 839, Sep. 14, 2026, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2026.133839.

[11] H. M. Kwan, G. Gao, F. Zhang, A. Gower, and D. Bull, "NVRC: Neural Video Representation Compression," in *Advances in Neural Information Processing Systems*, vol. 37, Curran Associates, Inc., 2024, pp. 132 440–132 462. DOI: 10.52202/079017-4210.

[12] T. Leguay, T. Ladune, P. Philippe, and O. Déforges, "Cool-chic video: Learned video coding with 800 parameters," in *2024 Data Compression Conference (DCC)*, Mar. 2024, pp. 23–32. DOI: 10.1109/DCC58796.2024.00010.

[13] L. Chen, Y. Zhao, J. C.-W. Chan, and S. G. Kong, "Histograms of oriented mosaic gradients for snapshot spectral image description," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 183, pp. 79–93, Jan. 1, 2022, ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2021.10.018.

[14] HotTracking2026, *Hyperspectral object tracking challenge 2026*, https://kaggle.com/competitions/hyperspectral-object-tracking-challenge-2026, Kaggle, 2026.

[15] S. Takamura and Y. Yashima, "Multiband video coding using H.264/AVC, MPEG-4 studio profile and JPEG 2000," in *Data Compression Conference, 2004. Proceedings. DCC 2004*, Mar. 2004, pp. 567–. DOI: 10.1109/DCC.2004.1281543.

[16] O. J. Pellicer-Valero, C. Aybar, and G. C. Valls. "Video Compression for Spatiotemporal Earth System Data." arXiv: 2506.19656 [cs], Accessed: Apr. 10, 2026. [Online]. Available: http://arxiv.org/abs/2506.19656, pre-published.

[17] F. Sippel, J. Seiler, and A. Kaup, "Synthetic hyperspectral array video database with applications to cross-spectral reconstruction and hyperspectral video coding," *JOSA A*, vol. 40, no. 3, pp. 479–491, Mar. 1, 2023, ISSN: 1520-8532. DOI: 10.1364/JOSAA.479552.

[18] Q. Du and J. E. Fowler, "Hyperspectral Image Compression Using JPEG2000 and Principal Component Analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201–205, Apr. 2007, ISSN: 1558-0571. DOI: 10.1109/LGRS.2006.888109.

[19] Z. Li et al., "A unified spatial-spectral-temporal network for hyperspectral object tracking," *Pattern Recognition*, vol. 174, p. 113 005, Jun. 1, 2026, ISSN: 0031-3203. DOI: 10.1016/j.patcog.2025.113005.